THE REVIEW OF SYMBOLIC LOGIC
Volume 4, Number 2, June 2011

IDENTITY AND INDISCERNIBILITY

JEFFREY KETLAND
Department of Philosophy, University of Edinburgh

Abstract. The notion of strict identity is sometimes given an explicit second-order definition: objects with all the same properties are identical. Here, a somewhat different problem is raised: *Under what conditions is the identity relation on the domain of a structure first-order definable*? A structure may have objects that are distinct, but indiscernible by the strongest means of discerning them given the language (the indiscernibility formula). Here a number of results concerning the indiscernibility formula, and the definability of identity, are collected and a number of applications discussed.

§1. Introduction. Informally, the Principle of Identity of Indiscernibles (PII) states that objects with "all the same properties"—that is, "indiscernible objects"—are identical. PII is usually formulated as the second-order principle:[1]

$$\forall x \forall y[\forall X(Xx \rightarrow Xy) \rightarrow x = y)]. \tag{1}$$

The antecedent $\forall X(Xx \rightarrow Xy)$ may informally be read as "any property of x is a property of y." This formula provides a *second-order definition of identity*, as follows:[2]

$$\forall x \forall y[x = y \leftrightarrow \forall X(Xx \rightarrow Xy)]. \tag{2}$$

To see how this works, suppose $\mathcal{M} = (D, R_1, \ldots, R_k)$ is a relational structure, but where the identity relation is *not* necessarily taken as primitive. Let $\mathcal{L}$ be the corresponding first-order language. We may expand $\mathcal{M}$ to a (standard, monadic) *second-order* structure $(\mathcal{M}, S)$, where S is $\mathcal{P}(D)$, and consider the language $\mathcal{L}_2$ obtained by adding (monadic) second-order quantifiers and variables. Then the formula $\forall X(Xx \rightarrow Xy)$ defines the identity relation on the structure $\mathcal{M}$. That is, for all $a, b \in D, a = b$ if and only if $(\mathcal{M}, S) \vDash \forall X(Xx \rightarrow Xy)[a, b]$. The left-to-right direction of this follows immediately from the fact that the formula $\forall X(Xx \rightarrow Xy)$ is reflexive. In the other direction, suppose a and b are distinct elements of D. Then $b \notin \{a\}$. Hence, since S contains *all* subsets of D, there is some $X \in S$ such that $a \in X$ and $b \notin X$. Thus, the formula $\forall X(Xx \rightarrow Xy)$ is false of (a, b). The crucial point for this argument is that, for each element $d \in D$, its unit set $\{d\}$ belongs to the range of second-order quantifiers.[3]

Received: August 25, 2007

[1] See, for example, Shapiro, 1991, p. 63; van Dalen, 1994, pp. 151–152; Manzano, 1996, pp. 2, 53–55

[2] Here we mean the notion of *strict* identity—that is, "$a = b$" means that a and b are one and the same object—and not some notion of similarity or qualitative indiscriminability.

[3] There has been a substantial debate concerning the status of PII. This includes a subliterature on alleged counterexamples to PII in metaphysics and physics (see Black, 1962; Cortes, 1976; French and Redhead, 1988; Saunders, 2002, 2003); and a subliterature on whether identity is, or should be, *definable* at all (see Frege, 1891; Savellos, 1990; Ketland, 2006); and a subliterature on the role of PII for *structuralist* views of mathematics. This is the "identity problem" for mathematical structuralism (see Burgess, 1999; Keränen, 2001; Ladyman, 2005; Ketland, 2006; Leitgeb & Ladyman, 2008; Shapiro, 2008).

© Association for Symbolic Logic, 2011

doi:10.1017/S1755020310000328

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

There are other second-order definitions. One may use the fact that the identity relation on a domain D is the *smallest reflexive binary relation* on D to define identity:

$$\forall x \forall y[x = y \leftrightarrow \forall R(\forall z Rzz \rightarrow Rxy)]. \tag{3}$$

To prove this, note first that the left-to-right direction follows from the fact that the formula $\forall R(\forall z Rzz \rightarrow Rxy)$ is reflexive. For the other direction, fix some domain D and let R be the diagonal on D: that is, $\{(a, a) \in D^2\}$. This is obviously reflexive. Suppose $a, b \in D$ and $a \neq b$. Thus, $(a, b) \notin R$. So, $\exists R(\forall z Rzz \wedge \neg Rab)$. And contraposition gives the result we want. Indeed, the above definition is equivalent to $\forall x \forall y[x \neq y \leftrightarrow \exists R(\forall z Rzz \wedge \neg Rxy)]$, itself equivalent to:

$$\forall x \forall y[x \neq y \leftrightarrow \exists R(\forall z \neg Rzz \wedge Rxy)] \tag{4}$$

So, x is distinct from y if and only if there is an irreflexive relation R such that Rxy. This, as we shall see, is intimately connected to Quine's notion of "weak discernibility," explained below.

Of course, the *identity relation* $=_D$ on a domain D can be defined as the *diagonal* of D. That is, $=_D$ is $\{(a, a) : a \in D\}$. However, if we examine the instance of comprehension needed to define this, the defining formula contains the identity predicate: that is, an arbitrary pair (a, b) is in $=_D$ if and only if $a \in D \wedge b \in D \wedge a = b$.

The definitions of identity (1) and (3) above are *second-order*.

However, we might feel some dissatisfaction with such second order definitions and instead ask whether identity is *first-order* definable. This is the central topic to be examined in this article: *Under what conditions is the identity relation in a structure first-order definable (without parameters)*?[4]

Another reasonable question concerns whether identity is *implicitly definable* by a set of first-order sentences. That is, if $\mathcal{L}$ is a first-order language without identity, containing some binary predicate symbol P, is there a set Δ of $\mathcal{L}$-sentences such that, for any model $\mathcal{M} \models \Delta$, we have that $P^{\mathcal{M}}$ is the identity relation on the domain of $\mathcal{M}$? The answer is quickly seen to be "no." For if $\mathcal{M}$ is a structure where $P^{\mathcal{M}}$ is the identity relation on $\mathcal{M}$, then there is an elementarily equivalent structure $\mathcal{M}^+$ where $P^{\mathcal{M}^+}$ is not the identity relation on its domain.[5]

The primary tool of investigation used here will be, for the kind of first-order language $\mathcal{L}$ under consideration, the *first-order indiscernibility formula*, written $x \approx_{\mathcal{L}} y$.[6] Throughout, let $\mathcal{M}$ be a relational structure of the form $(D, R_1, \ldots, R_k)$, with *finitely many* distinguished relations R_i, and the identity relation is not assumed as a primitive. The identity relation on D is denoted by "$=_{\mathcal{M}}$." Let $\mathcal{L}$ be the corresponding first-order language *without identity*, interpreted over $\mathcal{M}$ such that for each $i = 1$ to k, the primitive predicate symbol P_i denotes the distinguished relation R_i.[7] Let $(\mathcal{M}, =)$ be the result of expanding the structure $\mathcal{M}$ with the identity relation on D, and let $\mathcal{L}(=)$ be the result

[4] Some of the technical material below appears in Ketland (2006).

[5] This is pointed out also in Shapiro (2008). A proof is given below, Theorem 3.14.

[6] In model theory, a different notion of indiscernibility is studied: instead of a binary relation of indiscernibility, one defines the notions of a *set of indiscernibles* and a set of order-indiscernibles (see Hodges, 1997, pp. 152–153).

[7] The situation where one allows $\mathcal{M}$ to contain primitive *functions* (or constants) is left-open. However, $\mathcal{L}$ must contain only *finitely* many primitive predicate symbols. If $\mathcal{L}$ contains infinitely many predicate symbols, the corresponding indiscernibility formula $x \approx_{\mathcal{L}} y$ is no longer a formula of $\mathcal{L}$ itself. Rather, it is an infinitary conjunction.

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

of extending the language $\mathcal{L}$ by adding the identity symbol. Then, identity is first-order definable (without parameters) in a structure $\mathcal{M}$ if and only if there is an $\mathcal{L}$-formula $\varphi(x, y)$ such that $(\mathcal{M}, =) \vDash \forall x \forall y(x = y \leftrightarrow \varphi(x, y))$. This is equivalent to the demand that there exist a formula whose extension in $\mathcal{M}$ is indeed $=_{\mathcal{M}}$.

§2. Definitions. In the definitions below, "formula" always means "$\mathcal{L}$-formula." We shall always have $a, b \in D$; $\underline{d}$ is a finite sequence $(d_1, \ldots, d_n) \in D^n$, and $\underline{z}$ is a finite sequence of variables. If $f : D \to D$, then $f(\underline{d})$ is $(f(d_1), \ldots, f(d_n))$. "X" ranges over subsets of D and "R" ranges over subsets of D^n, for $n > 1$. If $\varphi(x)$ is a formula with exactly x free, then we use "$\varphi(a)$" as a convenient shorthand for "$a \in \varphi(\mathcal{M})$" (equivalently: "$\mathcal{M} \vDash \varphi(a)$"), where $\varphi(\mathcal{M})$ is the set defined by $\varphi(x)$; similarly, "$\varphi(a, b)$" is shorthand for "$(a, b) \in \varphi(\mathcal{M}^2)$," where $\varphi(\mathcal{M}^2)$ is the relation defined by $\varphi(x, y)$, where x and y are distinct variables; and so on.[8] If $\varphi(x)$ is a formula with y substitutable for x, then $\varphi(y)$ is the result of substituting y for *all* occurrences of x whenever x is free for y; and so on. And $\varphi_x(y)$ is the result of substituting y for *some or all* occurrences of x, wherever y is substitutable for x; and so on. We will sometimes write "$Rd_1d_2 \ldots d_n$" instead of "$(d_1, d_2, \ldots, d_n) \in R$."

On several occasions, W.V. Quine considers whether the notion of strict identity for a domain of individuals is definable or not, or whether some surrogate of identity, akin to indiscernibility, is sufficient for the relevant purposes.[9]

DEFINITION 2.1. *We say that a structure $\mathcal{M}$ is **Quinian** just in case $=_{\mathcal{M}}$ is first-order definable (without parameters) in $\mathcal{M}$. Otherwise, $\mathcal{M}$ is **non-Quinian**.*[10]

DEFINITION 2.2. *A formula $\varphi(x, y)$ is a **Leibniz formula** for $\mathcal{M}$ just in case,*

(a) $\mathcal{M} \vDash \forall x \varphi(x, x)$;

(b) For any formula $\theta(x, \underline{z})$, $\mathcal{M} \vDash \forall x \forall y(\varphi(x, y) \to \forall \underline{z}(\theta(x, \underline{z}) \to \theta_x(y, \underline{z})))$.

If $\varphi(x, y)$ is a Leibniz formula, then we say that it defines a **Leibniz relation** on $\mathcal{M}$. If a formula $\varphi(x, y)$ satisfies condition (b) above, we say it supports *substitutivity*.[11] Reasoning from $\varphi(a, b)$ and $\theta(a, \underline{d})$ to $\theta_a(b, \underline{d})$ is *reasoning by substitutivity* (of the formula $\varphi(x, y)$). Not all occurrences of a need be replaced by b.

DEFINITION 2.3. *Let $\mathbf{P}$ be a primitive n-ary predicate symbol ($n \geq 1$) of $\mathcal{L}$. Let $z_1, \ldots, z_{n-1}$ be a sequence of distinct variables, all distinct from x and y. Let $x \approx_{\mathbf{P}} y$ be the formula*

---

[8] This is a non-standard convention, but saves on repetitions of the context "$\mathcal{M} \vDash$". If the reader find this at all confusing in the proofs below, just replace occurrences of expressions like "$\phi(a)$", "$a \approx b$", etc., by "$\mathcal{M} \vDash \varphi(a)$", "$\mathcal{M} \vDash a \approx b$", and so on.

[9] See, for example, Quine, 1960, pp. 230–232, 1976, pp. 129–133, 1986, pp. 62–64.

[10] The adjective "Quinian" seems preferable to the commonly used "Quinean"; for, on the only occasion I am aware of Quine referring to himself this way, he writes: ". . . any more than there need be some peculiarly Quinian textural quality common to the protoplasm of my head and feet" (Quine, 1960, p. 171).

[11] As is well known, there are violations of substitutivity involving natural language predicates. For example, even though "Superman = Clark Kent" and "Lois believes that Superman can fly" are true, the result of substituting "Clark Kent" for "Superman" yields the sentence "Lois believes that Clark Kent can fly," which is false. It is a matter of contention what to say about such cases; but, in any case, here we shall ignore this kind of nonextensionality phenomenon, and concentrate entirely on extensional predicate logic.

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

$$\forall \underline{z}(\mathbf{P}xz_1 \ldots z_{n-1} \leftrightarrow \mathbf{P}yz_1 \ldots z_{n-1}) \wedge \forall \underline{z}(\mathbf{P}z_1xz_2 \ldots z_{n-1} \leftrightarrow \mathbf{P}z_1yz_2 \ldots z_{n-1}) \wedge \ldots \wedge \forall \underline{z}(\mathbf{P}z_1 \ldots z_{n-1}x \leftrightarrow \mathbf{P}z_1 \ldots z_{n-1}y).$$

*The **first-order indiscernibility formula** for $\mathcal{L}$, written $x \approx_{\mathcal{L}} y$, is the conjunction $\bigwedge\{x \approx_{\mathbf{P}} y : \mathbf{P}$ is a primitive predicate symbol of $\mathcal{L}\}$.*[12]

Henceforth, for ease of notation, we shall drop the subscript on $x \approx_{\mathcal{L}} y$.

Some familiar definitions:

DEFINITION 2.4. *Given a function $f : D \to D$ and a relation $R \subseteq D^k$, the image $f[R]$ is defined to be: $\{f(\underline{d}) : \underline{d} \in R\}$. We say that R is **invariant** under f if $f[R] = R$. A permutation $\pi : D \to D$ is called an **automorphism** of $\mathcal{M}$ if, for each distinguished relation R_i, $f[R_i] = R_i$. $\mathcal{M}$ is called **rigid** if its only automorphism is the identity mapping. $Aut(\mathcal{M})$ is the class of automorphisms of $\mathcal{M}$, and this is obviously a group under composition of permutations. The **transposition** $\pi_{ab} : D \to D$ is defined as follows: $\pi_{ab}(a) = b$ and $\pi_{ab}(b) = a$ and $\pi_{ab}(d) = d$ otherwise.*

Thus, a permutation $\pi : D \to D$ is an automorphism of $\mathcal{M}$ just when every distinguished relation R_i is invariant under π. (Note that the identity relation is trivially invariant under a permutation.) As is well known, if a relation $R \subseteq D^k$ is definable in a structure $\mathcal{M}$, then R is invariant under every $\pi \in Aut(\mathcal{M})$.

We introduce the main notions of indiscernibility as follows:[13]

DEFINITION 2.5.

*(1) a and b are **first-order indiscernible** in $\mathcal{M}$ iff $a \approx b$.*

*(2) a is **monadically indiscernible** from b in $\mathcal{M}$ iff there is no formula $\varphi(x)$ with exactly x free such that $\varphi(a)$ and $\neg\varphi(b)$.*

*(3) a is **polyadically indiscernible** from b in $\mathcal{M}$ iff, for any formula $\varphi(x, \underline{z})$, $\mathcal{M} \models \forall \underline{z}(\varphi(a, \underline{z}) \leftrightarrow \varphi(b, \underline{z}))$.*

*(4) a is **relatively indiscernible** from b in $\mathcal{M}$ iff there is no formula $\varphi(x, y)$ with exactly x and y free such that $\varphi(a, b)$ and $\neg\varphi(b, a)$.*

*(5) a is **weakly discernible** from b in $\mathcal{M}$ iff there is a formula $\varphi(x, y)$ such that $\neg\varphi(a, a)$ and $\varphi(a, b)$.*

*(6) a is **strongly indiscernible** from b in $\mathcal{M}$ iff a is not weakly discernible from b in $\mathcal{M}$*

*(7) a and b are **structurally indiscernible** in $\mathcal{M}$ iff there is a $\pi \in Aut(M)$ such that $b = \pi(a)$.*

Although we shall often omit explicit reference, each of these notions is defined relative to some structure $\mathcal{M}$.

DEFINITION 2.6. *We say that a set X **separates** a and b just in case ($a \in X$ iff $b \notin X$). And, where R is an n-ary relation and $\underline{d}$ an $(n-1)$-tuple, then we say that $(R, \underline{d})$ **separates** a and b just in case ($Rad_1 \ldots d_{n-1}$ iff $\neg Rbd_1 \ldots d_{n-1}$) or ($Rd_1ad_2 \ldots d_{n-1}$ iff*

[12] The first-order indiscernibility formula was first discussed by Hilbert & Bernays (1934, Vol 1, pp. 381), who noted that it plays the role of a surrogate for identity. The formula is also discussed by Quine (1960, p. 230, and 1986, pp. 63–64). As noted above, the relevant first-order language $\mathcal{L}$ must have only finitely many primitive symbols.

[13] Some of the notions below are borrowed from Quine, 1960, pp. 230–232, 1976, with some differences of terminology. Quine (1976) uses the term "discriminable." Instead, we follow some contemporary terminology and use "discernible."

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

$\neg Rd_1bd_2 \ldots d_{n-1}$) or ... or ($Rd_1 \ldots d_{n-1}a$ iff $\neg Rd_1 \ldots d_{n-1}b$). If it is clear what the intended relation is, we just say that the sequence $\underline{d}$ separates a and b. We say that a and b are ***separable*** *just in case some X, or some $(R, \underline{d})$, separates a and b. And, otherwise,* ***inseparable****. If the set, or relation, happens to be definable in a structure $\mathcal{M}$, we say that a and b are* ***definably separable*** *in $\mathcal{M}$.*

If elements a and b of a structure are distinct, then obviously the set $\{a\}$ (or, similarly, the set $\{b\}$) separates a and b. So, *objectively speaking*, distinct elements a and b are always separable. However, the separating set (or relation) might not be *definable* in some particular structure $\mathcal{M}$ under consideration.

Note finally that all of the notions of indiscernibility defined above are *model relative*. For example, we might consider a structure $\mathcal{M}$ containing the natural numbers 0 and 1, which are discernible in $\mathbb{N}$, but they may be *indiscernible* in this particular structure.

§3. Main results.

3.1. Properties of $x \approx y$**.** The second-order definition of identity says that a and b are identical just when they are not discernible by *any* property. However, when we restrict to a particular structure $\mathcal{M}$, and the usual first-order language for describing $\mathcal{M}$, then not *all* properties and relations need be *definable*. And, to speak loosely, we have $a \approx b$ exactly when a and b are not discerned by the properties and relations *definable* in the structure. Unsurprisingly then, we may have that a and b are indiscernible in a structure even though they are, in "reality," distinct elements of the domain.

Quite deliberately, the notion of *definable separability* is formulated so that each condition corresponds to a clause in the definition of $x \approx y$. It quickly follows that,

LEMMA 3.1. *$a \approx b$ if and only if a and b are not definably separable.*

The notion of a *Leibniz formula* encodes the basic formal properties of identity (i.e., reflexivity and substitutivity). It quickly follows that:

LEMMA 3.2. *If $\varphi(x, y)$ defines the identity relation, then $\varphi(x, y)$ is a Leibniz formula.*

And:

LEMMA 3.3. *If $\varphi(x, y)$ is a Leibniz formula, then the relation it defines is an equivalence relation.*

LEMMA 3.4. *Any definable reflexive subset of a Leibniz relation is also a Leibniz relation.*

Proof. Let $\varphi(x, y)$ be a Leibniz formula. Thus $\varphi(x, y)$ is reflexive. Suppose $\theta(x, y)$ is reflexive and $\mathcal{M} \vDash \forall x \forall y(\theta(x, y) \rightarrow \varphi(x, y))$. Suppose $\theta(x, y)$ is not a Leibniz formula. So, there is some formula $\psi(x, \underline{z})$ such that $\mathcal{M} \vDash \exists x \exists y(\theta(x, y) \wedge \exists \underline{z}(\psi(x, \underline{z}) \wedge \neg\psi(y, \underline{z})))$. Thus, there are $a, b, \underline{d} \in D$ such that $\theta(a, b)$ and $\psi(a, \underline{d})$ and $\neg\psi(b, \underline{d})$. Hence, by substitutivity of $\varphi(x, y)$, we have $\neg\varphi(a, b)$. Thus $\neg\theta(a, b)$. Contradiction. □

A proof by induction on the complexity of θ gives:

LEMMA 3.5. *For any formula $\theta(x, \underline{z})$, $\mathcal{M} \vDash \forall x \forall y(x \approx y \rightarrow \forall \underline{z}(\theta(x, \underline{z}) \rightarrow \theta_x(y, \underline{z})))$.*

Since $x \approx y$ is reflexive, this gives:

LEMMA 3.6. *The formula $x \approx y$ is a Leibniz formula.*

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

LEMMA 3.7. *If $\varphi(x, y)$ is a reflexive formula, then $\mathcal{M} \vDash \forall x \forall y(x \approx y \rightarrow \varphi(x, y))$.*

Proof. Suppose that $\varphi(x, y)$ is a reflexive formula and also $\mathcal{M} \vDash \exists x \exists y(x \approx y \wedge \neg\varphi(x, y))$. So, there are a, b such that $a \approx b$ but $\neg\varphi(a, b)$. Since $\varphi(x, y)$ is reflexive, $\varphi(a, a)$. But, by substitutivity, $\neg\varphi(a, a)$. Contradiction. □

Thus, the Leibniz relation $\approx_{\mathcal{M}}$ is a subset of any definable reflexive relation. In particular, this implies that $\approx_{\mathcal{M}}$ is a subset of any Leibniz relation on $\mathcal{M}$.

LEMMA 3.8. *If $\varphi(x, y)$ is a Leibniz formula, then $\mathcal{M} \vDash \forall x \forall y(\varphi(x, y) \rightarrow x \approx y)$.*

Proof. For suppose that $\varphi(x, y)$ is a Leibniz formula, and we have a, b such that $\varphi(a, b)$ and $\neg(a \approx b)$. Thus, since $\varphi(x, y)$ satisfies substitutivity, $\neg(a \approx a)$. Contradiction. □

Thus, if $\varphi(x, y)$ is a Leibniz formula, then the relation it defines is a subset of the relation $\approx_{\mathcal{M}}$.

Next, we may combine Lemmas 3.7 and 3.8 to give a uniqueness result:

THEOREM 3.9. *If $\varphi(x, y)$ is a Leibniz formula, then $\mathcal{M} \vDash \forall x \forall y(\varphi(x, y) \leftrightarrow x \approx y)$.*

Thus *any* Leibniz formula is equivalent over $\mathcal{M}$ to the first-order indiscernibility formula $x \approx y$. There is, up to equivalence in $\mathcal{M}$, exactly one Leibniz relation.[14] Of course, in a particular relational structure $\mathcal{M}$, a formula $\varphi(x, y)$ much simpler than $x \approx y$ may exist. However, it will be coextensive over $\mathcal{M}$ with $x \approx y$.

THEOREM 3.10. *If $\varphi(x, y)$ defines identity in $\mathcal{M}$, then $x \approx y$ defines identity too.*

Proof. It is obvious that if $a = b$, then $a \approx b$. Then suppose that $\varphi(x, y)$ defines identity in $\mathcal{M}$. So, $\varphi(a, b)$ implies $a = b$. And $\varphi(x, y)$ is a Leibniz formula. Thus, $a \approx b$ implies $\varphi(a, b)$. Hence, $a \approx b$ implies $a = b$. □

We have seen that any Leibniz relation is coextensive, in any structure $\mathcal{M}$, with the first-order indiscernibility relation $\approx_{\mathcal{M}}$.

The indiscernibility relation $\approx_{\mathcal{M}}$ is an equivalence relation. So, we can examine the quotient structure $\mathcal{M}/\approx_{\mathcal{M}}$, defined as follows.

DEFINITION 3.11. *For each element $a \in D$, we set $[a]$ to be the equivalence class $\{b \in D : b \approx_{\mathcal{M}} a\}$. The reduced domain D^- is $\{[a] : a \in D\}$. For an n-ary distinguished relation R in $\mathcal{M}$, we define the reduced relation R^- to be $\{([a_1], \ldots, [a_n]) : a_1, \ldots, a_n \in D\}$. Then $\mathcal{M}/\approx_{\mathcal{M}}$ is the reduced structure $(D^-, R_1{}^-, \ldots, R_k{}^-)$.*

THEOREM 3.12. $\mathcal{M} \equiv \mathcal{M}/\approx_{\mathcal{M}}$.

Proof. For ease of notation, let $\mathcal{M}^-$ be the quotient structure $\mathcal{M}/\approx_{\mathcal{M}}$. If $\sigma : Var(\mathcal{L}) \rightarrow \mathcal{M}$ is a $\mathcal{M}$-valuation, then the $\mathcal{M}^-$-valuation $\sigma^- : Var(\mathcal{L}) \rightarrow \mathcal{M}^-$ is defined as follows: for any variable x, $\sigma^-(x) = [\sigma(x)]$ (and thus $\sigma^-(x) \in D^-$). Then we prove:

> (*) For any $\mathcal{L}$-formula φ, any $\mathcal{M}$-valuation σ: $\mathcal{M}, \sigma \vDash \varphi$ if and only if $\mathcal{M}^-, \sigma^- \vDash \varphi$.

This is proved by induction. Let σ be an arbitrary $\mathcal{M}$-valuation.

[14] This result is mentioned and quickly proved in Quine (1962, p. 180).

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

(a) First, suppose φ is $\mathbf{P}x_1 \ldots x_n$. Then, $\mathcal{M}, \sigma \vDash \varphi$ if and only if $\mathcal{M}, \sigma \vDash \mathbf{P}x_1 \ldots x_n$, iff $(\sigma(x_1), \ldots, \sigma(x_n)) \in \mathbf{P}^{\mathcal{M}}$, iff $([\sigma(x_1)], \ldots, [\sigma(x_n)]) \in \mathbf{P}^{\mathcal{M}^-}$, iff $\mathcal{M}^-, \sigma^- \vDash \mathbf{P}x_1 \ldots x_n$, iff $\mathcal{M}^-, \sigma^- \vDash \varphi$, as required.
(b) Next let φ be of the form $\neg\theta$, where θ satisfies (*). Then, $\mathcal{M}, \sigma \vDash \varphi$ if and only if $\mathcal{M}, \sigma \vDash \neg\theta$, if and only if $\mathcal{M}, \sigma \nvDash \theta$, if and only if $\mathcal{M}^-, \sigma^- \nvDash \theta$, if and only if $\mathcal{M}^-, \sigma^- \vDash \neg\theta$, if and only $\mathcal{M}^-, \sigma^- \vDash \varphi$, as required.
(c) Let φ be of the form $\theta \rightarrow \psi$, where both θ and ψ satisfy (*). It quickly follows that φ satisfies (*) too, as required.
(d) Let φ be of the form $\forall x\theta$ where θ satisfies (*). Then, $\mathcal{M}, \sigma \vDash \varphi$ if and only if $\mathcal{M}, \sigma \vDash \forall x\theta$, if and only if, for all $a \in D$, $\mathcal{M}, \sigma(x|a) \vDash \theta$, if and only if, for all $a \in D, \mathcal{M}^-, \sigma^-(x|[a]) \vDash \theta$, if and only if $\mathcal{M}^-, \sigma^- \vDash \forall x\theta$, if and only $\mathcal{M}^-, \sigma^- \vDash \varphi$, as required.
Thus (*) is established. Elementary equivalence of $\mathcal{M}$ and $\mathcal{M}^-$ follows by taking φ to be a closed formula. □

LEMMA 3.13. *$[a] \approx_{\mathcal{M}^-} [b]$ if and only if $a \approx_{\mathcal{M}} b$.*

Proof. $[a] \approx_{\mathcal{M}^-} [b]$ if and only if $\mathcal{M}^-, \sigma^- \vDash x \approx y$, where $\sigma^-(x) = [a]$ and $\sigma^-(y) = [b]$. Let σ be any $\mathcal{M}$-valuation such that $\sigma(x) = a$ and $\sigma(y) = b$. By the result (*) above, we get $[a] \approx_{\mathcal{M}^-} [b]$ if and only if $\mathcal{M}, \sigma \vDash x \approx y$. So, $[a] \approx_{\mathcal{M}^-} [b]$ if and only if $a \approx_{\mathcal{M}} b$ □

THEOREM 3.14. *$\mathcal{M}/\approx_{\mathcal{M}}$ is Quinian.*

Proof. We need to show that, for any elements $c, d \in D^-$, $c \neq d$ implies $\neg(c \approx_{\mathcal{M}^-} d)$. Now, suppose that c and d are distinct elements of D^-. Since c and d are distinct, there are distinct equivalence classes $c = [a]$ and $d = [b]$ with $\neg(a \approx_{\mathcal{M}} b)$. By the previous lemma, $\neg([a] \approx_{\mathcal{M}^-} [b])$, and thus $\neg(c \approx_{\mathcal{M}^-} d)$, as required. □

We might call the structure $\mathcal{M}/\approx_{\mathcal{M}}$ the *Quinian quotient* of $\mathcal{M}$. In a sense, we can "invert" this construction by taking a structure (possibly Quinian) and adding "indiscernible" elements.

The idea is quite simple. For a nonempty relation R, take any object a in its field. Let b be any object not in the field of R. Then we force a and b to be "indiscernible" relative to a new relation R^+ by first defining a new relation R^+ to be the union $R \cup \{(b, d) : (a, d) \in R\} \cup \{(d, b) : (d, a) \in R\} \cup \{(b, b) : (a, a) \in R\}$. One may check that a and b are now indiscernible relative to the relation R^+. To apply this to a structure $\mathcal{M}$, we consider any object a in the domain, and take any object b not in the domain; for each distinguished relation R_i in $\mathcal{M}$, we define the new relations ${R_i}^+$. Then the new structure $\mathcal{M}^+$ has domain $dom(\mathcal{M}) \cup \{b\}$ and all the relations ${R_i}^+$ as primitive. Then we shall have $a \approx b$ in $\mathcal{M}^+$. A proof by induction shows that $\mathcal{M}$ and $\mathcal{M}^+$ are elementarily equivalent.

Suppose that D is a nonempty domain of objects. We might wonder if there is a set of formulas which *implicitly defines* the identity relation on D. In other words, is there a set of formulas all of whose models are isomorphic to $(D, =)$? The answer is no.

THEOREM 3.15. *Identity is not implicitly definable.*

Proof. Suppose that identity on D is implicitly definable. There is thus a first-order language $\mathcal{L}$ with a binary symbol $\mathbf{P}$, and a set Δ of $\mathcal{L}$-sentences such that $(D, =)$ is a model of Δ and, if $\mathcal{M} \vDash \Delta$, then $\mathcal{M}$ is isomorphic to $(D, =)$. The structure $(D, =)$ is a model of Δ and the relation $\mathbf{P}^{\mathcal{M}}$ is $\{(d, d) : d \in D\}$. Let a be any element of the domain and let b be an object not in D. Define the new relation $\mathbf{P}^{\mathcal{M}} \cup \{(a, b), (b, a), (b, b)\}$. Then, a and b

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

are now indiscernible. And the new structure $\mathcal{M}^+$, where this relation is the denotation of **P**, is elementarily equivalent to $\mathcal{M}$, but not isomorphic to $(D, =)$. Contradiction. □

This holds for domains of whatever cardinality: in general, we can add arbitrarily many elements to an original Quinian structure $\mathcal{M}$ to obtain an elementary equivalent non-Quinian structure (whose Quinian quotient is $\mathcal{M}$).

3.2. Notions of indiscernibility. There is a slightly different criterion for weak discernibility.

LEMMA 3.16. *a is weakly discernible from b if and only if there is a formula $\theta(x, y)$ such that $\mathcal{M} \vDash \forall x \neg\theta(x, x)$ and $\theta(a, b)$.*

Proof. The right-to-left direction is immediate. For the other direction, suppose that $\varphi(x, y)$ weakly discerns a and b. So, $\neg\varphi(a, a)$ and $\varphi(a, b)$. By substitutivity, $\neg(a \approx b)$. Let $\theta(x, y)$ be the formula $\varphi(x, y) \wedge \neg(x \approx y)$. Then, $\theta(d, d)$ iff $\varphi(d, d)$ and $\neg(d \approx d)$. Thus, $\neg\theta(d, d)$, for all d; and so $\theta(x, y)$ is irreflexive. Also, $\theta(a, b)$ iff $\varphi(a, b)$ and $\neg(a \approx b)$. And thus $\theta(a, b)$. □

Recall that a and b are *strongly indiscernibile* iff a and b are not weakly discernible.

THEOREM 3.17. *$a \approx b$ iff a and b are strongly indiscernible.*

Proof. For the right-to-left direction, suppose a and b are not weakly discernible, but $\neg(a \approx b)$. Then for any $\varphi(x, y)$, if $\neg\varphi(a, a)$, then $\neg\varphi(a, b)$. Let $\varphi(x, y)$ be $\neg(x \approx y)$. Then $\neg\varphi(a, a)$. Thus, $\neg\neg(a \approx b)$. Contradiction. For the left-to-right direction, suppose that a is weakly discernible from b and $a \approx b$. Hence, for some $\varphi(x, y)$, we have $\varphi(a, b)$ and $\neg\varphi(a, a)$. But by substitutivity, $\varphi(a, b)$ implies $\varphi(a, a)$. Contradiction. □

LEMMA 3.18. *$a \approx b$ if and only if a and b are polyadically indiscernible.*

Proof. Suppose that $a \approx b$ and suppose that, for some formula $\varphi(x, \underline{z})$, for some sequence $\underline{d}$, we have $\varphi(a, \underline{d})$. By substitutivity, we have $\varphi(b, \underline{d})$. Hence, $\varphi(a, \underline{d}) \rightarrow \varphi(b, \underline{d})$. Since $\underline{d}$ is arbitrary, we have $\mathcal{M} \vDash \forall \underline{z}(\varphi(a, \underline{z}) \rightarrow \varphi(b, \underline{z}))$. Thus, a and b are polyadically indiscernible.
For the other direction, suppose that a and b are polyadically discernible. Then, for all $\varphi(x, \underline{z})$, we have $\mathcal{M} \vDash \forall \underline{z}(\varphi(a, \underline{z}) \rightarrow \varphi(b, \underline{z}))$. So, for all d, if $a \approx d$, then $b \approx d$. So, if $a \approx a$, then $b \approx a$. But trivially $a \approx a$. So, $b \approx a$. And thus, $a \approx b$, as required. □

Thus polyadic indiscernibility corresponds exactly to first-order indiscernibility also.

LEMMA 3.19. *If a and b are first-order indiscernible, then a and b are relatively indiscernible.*

Proof. Suppose $\varphi(x, y)$ relatively discerns a and b. So, $\varphi(a, b)$ and $\neg\varphi(b, a)$. Suppose $a \approx b$. Then, $\varphi(a, a)$ and $\neg\varphi(a, a)$. Contradiction. □

LEMMA 3.20. *If a and b are relatively indiscernible, then a and b are monadically indiscernible.*

Proof. Suppose that the formula $\varphi(x)$ monadically discerns a and b, and so $\varphi(a)$ and $\neg\varphi(b)$. Let $\theta(x, y)$ be the formula $\varphi(x) \wedge \neg\varphi(y)$. Thus $\theta(a, b)$. And $\theta(b, a)$ just in case $\varphi(b)$ and $\neg\varphi(a)$, and thus $\neg\theta(b, a)$. Thus $\theta(x, y)$ relatively discerns a and b. □

Both of these inclusions are in fact proper. Given binary relations R_1 and R_2, say that R_1 is at least as strong as R_2 just in case $R_1 \subseteq R_2$. The (real) identity relation on a domain

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

D is the strongest reflexive binary relation on D. From what we already have, given a structure $\mathcal{M}$, the first-order indiscernibility relation $\approx_{\mathcal{M}}$ is the *strongest* indiscernibility notion which is *first-order definable* (without parameters) in $\mathcal{M}$. The other notions of indiscernibility may be properly weaker (i.e., proper supersets of the indiscernibility relation).

Consider the discrete linear order $(Z, <)$ of the integers. Then, any distinct pair is relatively discernible by the formula $x < y$, but no distinct pair is *monadically* discernible, since for any $z_1, z_2 \in Z$ there is a $\pi \in Aut(Z, <)$ such that $\pi(z_1) = z_2$.

For a toy example of a structure $\mathcal{M}$ with elements which are discernible but neither monadically nor relatively discernible, consider the simplest possible binary structure: just a set of objects with the identity relation as the sole primitive relation. That is, $\mathcal{M}$ has the form $(D, =)$. For example, let $D = \{0, 1\}$. Then, 0 and 1 are neither monadically nor relatively discernible, but obviously they are discernible.

3.3. Indiscernibility and automorphisms.

LEMMA 3.21. *If a and b are structurally indiscernible in $\mathcal{M}$, then a and b are monadically indiscernible.*

Proof. Suppose that a and b are structurally indiscernible. So, we have $\pi \in Aut(\mathcal{M})$ such that $\pi(a) = b$. Suppose that $\varphi(a)$. Since the set defined by $\varphi(x)$ is invariant under any automorphism $\pi \in Aut(\mathcal{M})$, we have $\varphi(a)$ iff $\varphi(\pi(a))$. Thus, $\varphi(b)$. So, a and b are monadically indiscernible. □

Similarly,

LEMMA 3.22. *If there is some $\pi \in Aut(\mathcal{M})$ such that $\pi(a) = b$ and $\pi(b) = a$, then a and b are relatively indiscernible.*

Proof. Suppose that a and b are relatively discerned by $\varphi(x, y)$. Thus, $\varphi(a, b)$ and $\neg\varphi(b, a)$. Suppose we have $\pi \in Aut(\mathcal{M})$ such that $b = \pi(a)$ and $a = \pi(b)$. Since $\pi \in Aut(\mathcal{M})$, we have $\varphi(a, b)$ iff $\varphi(\pi(a), \pi(b))$ iff $\varphi(b, a)$. Thus, $\varphi(b, a)$. Contradiction. □

The next result is quite useful:

THEOREM 3.23. *If $a \approx b$ then $\pi_{ab} \in Aut(\mathcal{M})$.*

Proof. Suppose $a \approx b$. Then, for any distinguished set X or relation R, a and b are indiscernible. For ease of notation, let π be the transposition π_{ab}. We aim to show that π is an automorphism of $\mathcal{M}$. First, we show that if a and b are not discernible by a set X, then $\pi[X] = X$ (where $\pi[X]$ is the image of X under π). For suppose $a \in X$ iff $b \in X$. Then $\pi(b) \in X$ iff $\pi(a) \in X$. Since π also leaves all other elements invariant, it follows that $\pi[X] = X$. We similarly show that if a and b are not discernible by a binary relation R, then $\pi[R] = [R]$. For suppose, for all d, $(a, d) \in R$ iff $(b, d) \in R$ and $(d, a) \in R$ iff $(d, b) \in R$. It then follows (running through possible cases) that, for all (d_1, d_2), $(\pi(d_1), \pi(d_2)) \in R$ iff $(d_1, d_2) \in R$. Thus, $\pi[R] = R$. And so on for all relations of higher arity. So, every distinguished set or relation is invariant under π. So, π is an automorphism. □

We can sometimes use this result to show that a distinct pair of elements in a given structure $\mathcal{M}$ are discernible. However, the converse of this result is *not* true. π_{ab} may be an automorphism even though a and b are first-order discernible. For example, let $D = \{0, 1\}$ and consider the structure $(D, =)$. Obviously, π_{01} is an automorphism, but 0 and 1 are

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

discernible. So, in general, if π_{ab} is an automorphism, it will follow that a and b are relatively indiscernible (by Lemma 3.22); but it need not follow that $a \approx b$. However, the following is a near converse:

LEMMA 3.24. *Suppose some* $(R, \underline{d})$ *separates* a *and* b *and none of the* d_i *is either* a *or* b. *Then* π_{ab} *is not an automorphism.*

Proof. Suppose a and b are separated by $(R, \underline{d})$, and none of the d_i is either a or b. So, either $Rad_1 \dots d_n$ iff $\neg Rbd_1 \dots d_n$, or $\dots$ or $Rd_1 \dots a$ iff $\neg Rd_1 \dots d_n b$. Since none of the d_i is either a or b, $\pi_{ab}(d_i) = d_i$. Suppose π_{ab} is an automorphism. So, $Rad_1 \dots d_n$ iff $Rbd_1 \dots d_n$, and $\dots$ $Rd_1 \dots d_n a$ iff $Rd_1 \dots d_n b$. This contradicts the claim that $(R, \underline{d})$ separates a and b. □

The converse of Theorem 3.23 *is* true for monadic structures. Suppose that $\mathcal{M}$ is a monadic structure and π_{ab} is an automorphism. Thus, no distinguished set X_i separates a and b. So, $a \in X_i$ iff $b \in X_i$. Hence, $a \in X_i$ iff $\pi_{ab}(a) \in X_i$. So, $a \approx b$.

3.4. Criteria for the definability of identity. Recall that $\mathcal{M}$ is called *Quinian* just in case $=_{\mathcal{M}}$ is first-order definable (without parameters) in $\mathcal{M}$. Quine himself drew attention to *non*-Quinian structures:

> It may happen that the objects intended as values of the variables of quantification are not completely distinguishable from one another by the four predicates. When this happens, [the indiscernibility formula] fails to define genuine identity. Still such failure remains unobservable from within the language (Quine, 1986, p. 63).

We proceed now to identify some criteria for a structure to be Quinian.

THEOREM 3.25. $\mathcal{M}$ *is Quinian if and only if every distinct pair* a, b *in* $\mathcal{M}$ *is definably separable.*

Proof. For the left-to-right direction, suppose $=_{\mathcal{M}}$ is definable. Then, by Theorem 3.10, $x \approx y$ defines $=_{\mathcal{M}}$. In particular, $a \approx b$ implies $a = b$. So, if a and b are distinct, then they are discernible, and thus separable, and definably so. In the other direction, suppose that every distinct pair is separable, but identity is not definable. Thus, $x \approx y$ does not define identity. Hence, for some $a \neq b$ we have $a \approx b$. But a and b are separable, thus $\neg(a \approx b)$. Contradiction. □

Let T be $Th(\mathcal{M})$. Let $T^{=}$ be $Th(\mathcal{M}, =)$. Clearly:

LEMMA 3.26. $\mathcal{M}$ *is Quinian if and only if* $T^{=} \vdash \forall x \forall y (x = y \leftrightarrow x \approx y)$.

Proof. The right-to-left direction is immediate. For the left-to-right direction, we use Theorem 3.10. □

We may use the Beth definability theorem to show the following:

THEOREM 3.27. $\mathcal{M}$ *is Quinian iff, for all* $R \subseteq D^2$, *if* $(\mathcal{M}, R) \models T^{=}$ *then* R *is* $=_{\mathcal{M}}$.

Proof. Suppose $=_{\mathcal{M}}$ is definable in $\mathcal{M}$ and $(\mathcal{M}, R) \models T^{=}$. Then, $(\mathcal{M}, R) \models \forall x \forall y (x = y \leftrightarrow x \approx y)$. Thus, R is identical to $\approx_{\mathcal{M}}$. But $\approx$ defines $=_{\mathcal{M}}$, and thus R is $=_{\mathcal{M}}$. For the converse, suppose that for all $R \subseteq D^2$, if $(\mathcal{M}, R) \models T^{=}$, then R is $=_{\mathcal{M}}$. It follows that $=_{\mathcal{M}}$ is *implicitly defined* in $T^{=}$. By the Beth Definability Theorem, $T^{=}$ contains an explicit definition of $=_{\mathcal{M}}$. □

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

The following two theorems set out some general and useful criteria for the definability of identity.

THEOREM 3.28. *Each of the three conditions below is both necessary and sufficient for* $\mathcal{M}$ *to be Quinian:*

(a) for all $a, b \in D$, $a \approx b$ *implies* $a = b$;

(b) for all $R \subseteq D^2$, *if* $(\mathcal{M}, R) \models T^{=}$ *then* R *is the identity relation on* D.

(c) some surjective total function is definable in $\mathcal{M}$.

Proof. (a) and (b) have been established above.
(c) Necessity is trivial. For if identity is definable, then it is definable by $x \approx y$, and this formula defines the identity *function*, which is surjective and total! For sufficiency, suppose that, for some $n \geq 1$, the formula $\varphi(\underline{z}, w)$ defines some surjective total function $f : D^n \to D$, where $\underline{z}$ is an ordered n-tuple of distinct variables. So, $\varphi(\underline{d}, a)$ if and only if $f(\underline{d}) = a$. Let $\theta(x, y)$ be the formula $\exists \underline{z}(\varphi(\underline{z}, x) \wedge \varphi(\underline{z}, y))$. Then $\theta(x, y)$ defines identity. For $\theta(a, a)$ if and only if $\exists \underline{z}\varphi(\underline{z}, a)$, iff, for some $\underline{d}$, $f(\underline{d}) = a$, and by surjectivity, this is so. And suppose $\exists \underline{z}(\varphi(\underline{z}, a) \wedge \varphi(\underline{z}, b))$. Then, by functionality, $a = b$. □

THEOREM 3.29. *Each of the three conditions below is sufficient (but not necessary) for* $\mathcal{M}$ *to be Quinian:*

(a) $\mathcal{M}$ *is rigid;*

(b) no nontrivial transposition π_{ab} *is an automorphism of* $\mathcal{M}$;

(c) a strict linear order is definable in $\mathcal{M}$.

Proof. (a) Suppose identity is not definable. Then we have $a \approx b$ for some $a \neq b$. Then, by Theorem 3.23, π_{ab} is an automorphism. But π_{ab} is nontrivial. So, $\mathcal{M}$ is not rigid.
(b) Suppose identity is not definable. Then we have $a \approx b$ for some $a \neq b$. Thus, π_{ab} is an automorphism.
(c) Suppose that $\varphi(x, y)$ defines a strict linear order. Then $\neg\varphi(x, y) \wedge \neg\varphi(y, x)$ defines identity. □

None of these three conditions is, in general, necessary.[15] To see the nonnecessity of (a)–(c), consider the structure $(D, =)$, with D any set with cardinality greater than 1. For definiteness, suppose $D = \{0, 1\}$. Trivially $(D, =)$ is Quinian, but every permutation of the domain is an automorphism (and thus any nontrivial transposition is an automorphism). Note that 0 and 1 are not *relatively* discernible in $(D, =)$. If a linear order were definable, then 0 and 1 would be relatively discernible. So, a linear order is not definable in $(D, =)$. So, $(D, =)$ is a Quinian structure in which a linear order is not definable. A slightly fancier example is the complex field $\mathbb{C}$, thought of as a relational structure. Although $\mathbb{C}$ is Quinian (since a surjective total function is definable), $\mathbb{C}$ is nonrigid and no linear order is definable in $\mathbb{C}$.

§4. Some applications.

4.1. Conservation of identity. It is unsurprising that extending a theory T in $\mathcal{L}$ with the usual axioms for identity (reflexivity and some version of substitutivity) results in a

[15] Condition (a) is necessary in the somewhat uninteresting case of monadic structures. If $\mathcal{M} = (D, X_1, \ldots, X_n)$ is a *monadic* structure, then identity is definable in $\mathcal{M}$ if and only if $\mathcal{M}$ is rigid.

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

conservative extension. This may be seen by the fact that any model $\mathcal{M}$ of such a theory T can be (trivially) expanded to the model $(\mathcal{M}, =)$ of the theory with the axioms of identity added. It is worth noting here that a proof-theoretic proof of this result can also be obtained, using the properties of the first-order indiscernibility formula $x \approx y$.

First note that the reflexivity, and substitutivity, of the first-order indiscernibility formula are provable in logic alone (without identity):

LEMMA 4.1. $\vdash \forall x(x \approx x)$.

LEMMA 4.2. *For any* $\varphi(x, \underline{z})$, $\vdash \forall x \forall y(x \approx y \rightarrow \forall \underline{z}(\varphi(x, \underline{z}) \rightarrow \varphi_x(y, \underline{z})))$.

Second, for a formula φ containing the identity predicate, we can define a formula $\varphi^{\approx}$, with the same free variables, obtained by substituting the first-order indiscernibility formula for occurrences of the identity predicate (i.e., any occurrence of $x = y$ is replaced by $x \approx y$).

Next, suppose that T in $\mathcal{L}$ is a theory in a language lacking the identity symbol, and $T^{=}$ is obtained by adding the usual axioms for identity in the language $\mathcal{L}(=)$. Then, we have:

THEOREM 4.3. *For any* $\mathcal{L}$*-formula* φ*, if* $T^{=} \vdash \varphi$ *then* $T \vdash \varphi$.

Proof. Consider a derivation $(\varphi_1, \ldots, \varphi_n)$ in $T^{=}$ of an $\mathcal{L}$-formula φ. Replace each φ_i by the corresponding $=$-free $\mathcal{L}$-formula $\varphi_i^{\approx}$, and for each axiom of identity φ that appears, insert the corresponding subderivation in predicate logic of $\varphi^{\approx}$. The result is then a derivation in T of φ. □

4.2. Quinian and non-Quinian structures Let D be $\{0, 1\}$ and let R be the relation $\{(0, 0), (1, 1), (0, 1), (1, 0)\}$. We can show that (D, R) is non-Quinian. We have that, for all $d_1, d_2 \in \{0, 1\}$, Rd_1d_2. So, for all d, $R0d$ iff $R1d$ and $Rd0$ iff $Rd1$. Thus, no element d definably separates 0 and 1. And thus identity is not definable. More generally, let $\mathcal{M} = (D, R)$, where D is any set with $|D| > 1$ and $R = \{(d_1, d_2) : d_1, d_2 \in D\}$. Then $\mathcal{M}$ is non-Quinian. For suppose that identity is definable. Then some element d would definably separate some pair $a, b \in D$. Thus, either Rda iff $\neg Rdb$ or Rad iff $\neg Rbd$. But both cases are impossible. Notice that every permutation of the domain of $\mathcal{M}$ is an automorphism.

Although we have required that $\mathcal{M}$ be a relational structure of the form $(D, R_1, \ldots, R_k)$, we *of course* allow that one of the relations R_i may, in fact, extensionally *be* a function on D. And, as we have seen, if one of relations these R_i is a surjective total function, then identity is definable in $\mathcal{M}$. So, consider algebraic structures regarded as *relational* structures, but wherein identity is *not* treated as primitive. Theorem 3.28(c) shows, for example, that any group $\mathcal{G} = (G, \cdot)$ is Quinian, for the primitive ternary relation $\cdot$ is a surjective total function on the domain G (i.e., for all $a \in G$, there exist $c, d \in G$ such that $c \cdot d = a$).

Similarly, Theorem 3.29 shows that any strict *linear* order $(D, <)$ is Quinian. For example, the formula $\neg(x < y) \wedge \neg(y < x)$ defines identity. Similarly, identity is definable in any ordinal when it is thought of as a relational structure. However, in a strict partial ordering $(D, <)$, one may have two incomparable elements a, b such that $a \not< b$ and $b \not< a$, but a and b have all the same smaller elements and all the same larger elements. In this case, a and b are nonseparable, and thus indiscernible. Then, $(D, <)$ is non-Quinian.

Turning to geometry, consider $\mathbb{R}^3$ with its natural Euclidean geometric structure. More precisely, let $\mathbb{E}^3$ be the structure with domain $\mathbb{R}^3$ and distinguished betweenness relation $Bet(\mathbb{R}^3)$ and congruence relation $Cong(\mathbb{R}^3)$. The structure $\mathbb{E}^3$ can be regarded as the unique-up-to-isomorphism structure characterized by Hilbert's second-order system of

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

axioms for geometry.[16] The space $\mathbb{E}^3$ has a great many symmetries, best understood as coordinate transformations $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ (translations, rotations, reflections, inversions, and dilations), and thus is nonrigid. But it is Quinian. Indeed, identity is very simply definable, since one of the *axioms* of geometry is $\forall x \forall y(Bxyx \leftrightarrow x = y)$.[17]

4.3. A topological application. The notion of separability defined above hints of a connection with the usual *topological* notions of separability. Let $\mathcal{S} = (D, \mathcal{U})$ be a topological space, with $\mathcal{U} \subseteq \mathcal{P}(D)$ satisfying usual conditions. Then we say that elements a and b are *topologically distinguishable* in $\mathcal{S}$ just in case there is an open set $O \in \mathcal{U}$ such that $a \in O$ and $b \notin O$. We say that $\mathcal{S}$ is T_0 (or *Kolmogorov*) if and only if any pair of distinct elements $a, b \in D$ are topologically distinguishable. The usual example of a T_0 space is the standard topology of the reals, generated from the open intervals as basis: since if r_1, r_2 are distinct reals, then there is an open interval (s, t) such that $r_1 \in (s, t)$ and $r_2 \notin (s, t)$.

To see the connection with our notion of separability, define an associated binary relational structure $\mathcal{M}_\mathcal{S} = (D \cup \mathcal{U}, \in\restriction_{D \cup \mathcal{U}})$. We will show that, with a certain side condition, the T_0 separability of $\mathcal{S}$ implies the definable separability of $\mathcal{M}_\mathcal{S}$, and thus the definability of identity in $\mathcal{M}_\mathcal{S}$ by the indiscernibility formula $x \approx y$ (i.e., the formula $\forall z[(z \in x \leftrightarrow z \in x) \wedge (x \in z \leftrightarrow y \in z)]$).

However, a small wrinkle appears here because the domain of our structure is the union $D \cup \mathcal{U}$, and we need to ensure that the base set D and the topology $\mathcal{U}$ are disjoint. This is achieved by imposing the following conditions:

(i) $\emptyset \notin D$

(ii) $\forall x, y \in D(x \notin y)$.

Then, we obtain

LEMMA 4.4. *Let D be a set. Then, if conditions (i) and (ii) hold, no element $d \in D$ coincides with any subset $X \subseteq D$.*

Proof. Suppose $d \in D$ and $X \subseteq D$. If X is $\emptyset$, then d and X are clearly distinct. If X is not $\emptyset$, then $a \in X$, for some a. So, $a \in D$. Hence, $a \notin d$. So, d and X are distinct. □

This gives the following result:

THEOREM 4.5. *Suppose that $\mathcal{S} = (D, \mathcal{U})$ is a topological space, and that D satisfies conditions (i) and (ii) above. Then $\mathcal{M}_\mathcal{S}$ is Quinian if and only if $\mathcal{S}$ is T_0.*

Proof. For the left-to-right direction, suppose that identity is definable in $\mathcal{M}_\mathcal{S}$. Then $a \approx b$ implies $a = b$. In particular, for any elements $a, b \in D$, $a \approx b$ implies $a = b$. Suppose a and b are distinct. Then, $\neg(a \approx b)$. Thus, there is some $d \in D \cup \mathcal{U}$ such that $a \in d$ iff $b \notin d$ or $d \in a$ iff $d \notin b$. In the latter case, it follows that either a or b has an

[16] To be more exact—a version of Hilbert's axiom system in which only points are treated as basic. One may formulate a first-order theory E_3 of this structure and prove a representation theorem: $\mathcal{M} \vDash E_3$ iff $\mathcal{M}$ is isomorphic to $(\mathbb{F}^3, Bet(\mathbb{F}^3), Cong(\mathbb{F}^3))$, where $\mathbb{F}$ is a real-closed field. The theory E_3 is complete and decidable. See Tarski (1959, pp. 169–170) for more details.

[17] See Tarski (1959, p. 166). It needs to be stressed that identity is taken to be a *primitive* in the language of this theory, and one of the axioms is $\forall x \forall y(Bxyx \leftrightarrow x = y)$. If identity is not taken to be a primitive, and one tries to reformulate the theory with just $Bxyz$ and $Cxyzw$ as primitives (and then, say, introducing $x = y$ as defined by $Bxyx$), then the models of the theory are no longer what was intended.

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

element in D, which is impossible since, by (ii), $\forall x, y \in D(x \notin y)$. Suppose the former case. It follows that d is nonempty, and thus is some $O \in \mathcal{U}$. Thus, $a \in O$ iff $b \notin O$, and thus $\mathcal{S}$ is T_0.

For the right-to-left direction, suppose $\mathcal{S}$ is T_0. Hence, for any distinct $a, b \in D$, there is some open set $O \in \mathcal{U}$ such that $a \in O$ and $b \notin O$. Hence, for any distinct $a, b \in D$, there is an element $d \in D \cup \mathcal{U}$ such that $a \in d$ iff $b \notin d$. Hence, $\neg(a \approx b)$. Furthermore, each distinct pair in $\mathcal{U}$ is discernible using $\in$, by extensionality. Finally, by the above lemma, if $d \in D$ is distinct from $O \in U$, then they are separable. Thus, identity is definable, by the formula $x \approx y$. □

This theorem provides a general method for constructing non-Quinian structures, in which identity is *not* definable. Begin with any non-T_0 topological space $\mathcal{S}$ satisfying the side condition, and then the corresponding relational structure $\mathcal{M}_{\mathcal{S}}$ is non-Quinian.

For a simple example, let $D = \{0, 1\}$ and let the topology $\mathcal{U}$ be the trivial one: $\{\emptyset, \{0, 1\}\}$. Clearly, $\mathcal{S}$ is not T_0: the elements 0 and 1 are topologically indistinguishable. Then, by the theorem above, the relational structure $(D \cup U, \in\restriction_{D \cup \mathcal{U}})$ is non-Quinian. This 4-element structure is isomorphic to the binary structure (D', R) where D' is $\{0, 1, 2, 3\}$ and R is $\{(0, 3), (1, 3)\}$ (the element 2 corresponds to the empty set $\emptyset$ and the element 3 corresponds to $\{0, 1\}$). It is clear that 0 and 1 are first-order indiscernible, for there is no element d which separates them.

§5. Acknowledgment. This article is based on a 2005 manuscript discussing logical issues surrounding identity and indiscernibility, and given as a talk at Bristol University in 2006 and 2009. I am grateful for comments from the late Torkel Franzén, and from the audiences at Bristol.

BIBLIOGRAPHY

Black, M. (1962). The identity of indiscernibles. *Mind*, **61**, 153–164.

Burgess, J. P. (1999). Review of Shapiro 1997. *Notre Dame Journal of Formal Logic*, **40**, 283–291.

Cortes, A. (1976). Leibniz's principle of identity of indiscernibles: A false principle. *Philosophy of Science*, **45**, 466–470.

Frege, G. (1891). Review of E. Husserl 1891, *Philosophie der Arithmetik*. Extracts reprinted in Geach & Black (eds.) 1980; Translations from the Philosophical Writings of Gottlob Frege (third edition). New Jersey, NJ: Barnes and Noble.

French, S., & Redhead, M. (1988). Quantum physics and the identity of indiscernibles. *British Journal for the Philosophy of Science*, **39**, 233–246.

Hilbert, D., & Bernays, P. (1934). *Grundlagen der Mathematik*, Vol. 1. Berlin: Springer.

Hodges, W. (1997). *A Shorter Model Theory*. Cambridge, MA: Cambridge University Press.

Keränen, J. (2001). The identity problem for realist structuralism. *Philosophia Mathematica*, **3**, 308–330.

Ketland, J. (2006). Structuralism and the identity of indiscernibles. *Analysis*, **66**, 303–315.

Ladyman, J. (2005). Mathematical structuralism and the identity of indiscernibles. *Analysis*, **62**, 218–221.

Leitgeb, H., & Ladyman, J. (2008). Criteria of identity and structuralist ontology. *Philosophia Mathematica*, **16**, 388–396.

Manzano, M. (1996). *Extensions of First-Order Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge, MA: Cambridge University Press.

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press

Quine, W. V. (1960). *Word and Object*. Cambridge, MA: MIT Press.
Quine, W. V. (1962). Reply to Professor Marcus. *Synthese*, **20**. Page references are to the reprint in W. V. Quine 1976, *The Ways of Paradox* (revised edition). Cambridge, MA: Harvard University Press.
Quine, W. V. (1976). Grades of Discriminability. *Journal of Philosopy*, **73**. Reprinted in W. V. Quine 1981, *Theories and Things* (Cambridge, MA: Harvard University Press), pp. 129–133.
Quine, W. V. (1986). *Philosophy of Logic* (second edition). Cambridge, MA: Harvard University Press.
Saunders, S. (2002). Indiscernibles, general covariance and other symmetries: The case for non-eliminativist relationism. In Ashtekar, A., Howard, D., Renn, J., Sarkar, S., & Shimony, A., editors. *Revisiting the Foundations of Relativistic Physics: Festschrift in Honour of John Stachel*. Dordrecht, The Netherlands: Kluwer.
Saunders, S. (2003). Physics and Leibniz's principles. In Castellani, E., and Brading, K., editors. *Symmetries in Physics: Philosophical Reflections*, Cambridge, MA: Cambridge University Press.
Savellos, E. (1990). On defining identity. *Notre Dame Journal of Formal Logic*, **31**, 476–484.
Shapiro, S. (1991). *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford: Oxford University Press.
Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press.
Shapiro, S. (2008). Identity, indiscernibility and *ante rem* structuralism: The tale of i and $-i$. *Philosophia Mathematica*, **16**, 285–309.
Tarski, A. (1959). What is elementary geometry? In Henkin, L., Suppes, P., and Tarski, A., editors. *The Axiomatic Method*, Amsterdam, The Netherlands: North Holland, pp. 16–29. Page references are to the reprint in J. Hintikka (ed.) 1968, *Philosophy of Mathematics*. Oxford: Oxford University Press.
van Dalen, D. (1994). *Logic and Structure*. Berlin: Springer.

JEFFREY KETLAND
DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF EDINBURGH
EDINBURGH
UNITED KINGDOM
E-mail: jeffrey.ketland@ed.ac.uk

https://doi.org/10.1017/S1755020310000328 Published online by Cambridge University Press